

A00829022_MomentoDeRetroalimentacion_Modulo5ProcesamientoDeDatosMultivariados_PortafolioImplementacion

Josue Salvador Cano Martinez

2022-10-25

Cargar datos, eliminar tanto columna de id como datos no cuantitativos

```
D=read.csv("mercurio.csv")
D$X1 <- NULL
D$X2 <- NULL
head(D)
```

		X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
##	1	5.9	6.1	3.0	0.7	1.23	5	0.85	1.43	1.53	1
##	2	3.5	5.1	1.9	3.2	1.33	7	0.92	1.90	1.33	0
##	3	116.0	9.1	44.1	128.3	0.04	6	0.04	0.06	0.04	0
##	4	39.4	6.9	16.4	3.5	0.44	12	0.13	0.84	0.44	0
##	5	2.5	4.6	2.9	1.8	1.20	12	0.69	1.50	1.33	1
##	6	19.6	7.3	4.5	44.1	0.27	14	0.04	0.48	0.25	1

Mardia's Test

H_0 Las variables siguen una distribución normal multivariable

H_1 Las variables no siguen una distribución normal multivariable

```
library(QuantPsyc)

## Loading required package: boot
## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
## Loading required package: purrr
## Loading required package: MASS
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##      select
##
## Attaching package: 'QuantPsyc'
## The following object is masked from 'package:base':
##
##      norm
mult.norm(D)$mult.test

##           Beta-hat      kappa      p-val
## Skewness  53.74505 474.747945 0.0000000000
## Kurtosis 135.31273   3.597949 0.0003207365
```

Debido a que los dos p-values son menores a 0.05, se niega la hipótesis H_0 , no se tiene evidencia para decir que las variables del set de datos siguen una distribución multivariante.

Anderson-Darling Test

H_0 Los datos siguen una distribución normal

H_1 Los datos no siguen una distribución normal

```
library(nortest)
ad.test(D$X3)

##
## Anderson-Darling normality test
##
## data: D$X3
## A = 3.6725, p-value = 2.706e-09

ad.test(D$X4)

##
## Anderson-Darling normality test
##
## data: D$X4
## A = 0.34956, p-value = 0.4611

ad.test(D$X5)
```

```
##
## Anderson-Darling normality test
##
## data: D$X5
## A = 4.051, p-value = 3.193e-10
ad.test(D$X6)

##
## Anderson-Darling normality test
##
## data: D$X6
## A = 5.4286, p-value = 1.4e-13
ad.test(D$X7)

##
## Anderson-Darling normality test
##
## data: D$X7
## A = 0.92528, p-value = 0.0174
ad.test(D$X8)

##
## Anderson-Darling normality test
##
## data: D$X8
## A = 8.6943, p-value < 2.2e-16
ad.test(D$X9)

##
## Anderson-Darling normality test
##
## data: D$X9
## A = 1.977, p-value = 4.161e-05
ad.test(D$X10)

##
## Anderson-Darling normality test
##
## data: D$X10
## A = 0.65847, p-value = 0.08099
ad.test(D$X11)

##
## Anderson-Darling normality test
##
## data: D$X11
## A = 1.0469, p-value = 0.008637
```

```
ad.test(D$X12)

##
## Anderson-Darling normality test
##
## data: D$X12
## A = 14.335, p-value < 2.2e-16
```

Tomando como consideración un valor de significancia p-value igual a 0.05, Las variables que niegan la hipótesis H_0 y que por tanto tienden a tener una distribución normal son: X4 y X10

Mardia's Test a variables que presentan normalidad

H_0 Las variables siguen una distribución normal multivariable

H_1 Las variables no siguen una distribución normal multivariable

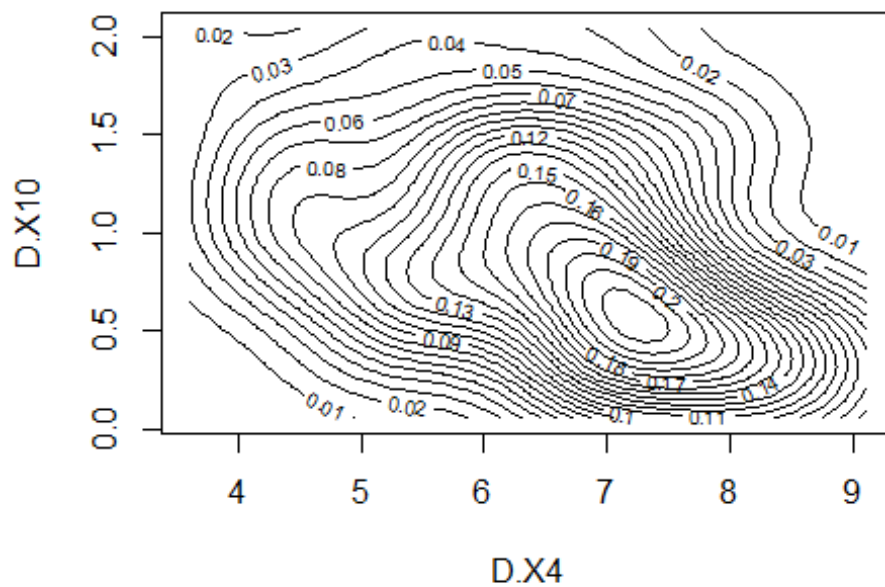
```
normD <- data.frame(D$X4, D$X10)
library(QuantPsyc)
mult.norm(normD)$mult.test
```

```
##           Beta-hat      kappa      p-val
## Skewness 0.6991004  6.175387 0.1864276
## Kurtosis 6.7602297 -1.128208 0.2592321
```

Debido a que los valores de p-value son mayores a 0.05, se puede entender un nivel de significancia suficiente para negar la hipótesis H_1 y deducir que las variables X4 y X5 siguen una distribución normal multivariable.

Gráfica de contorno

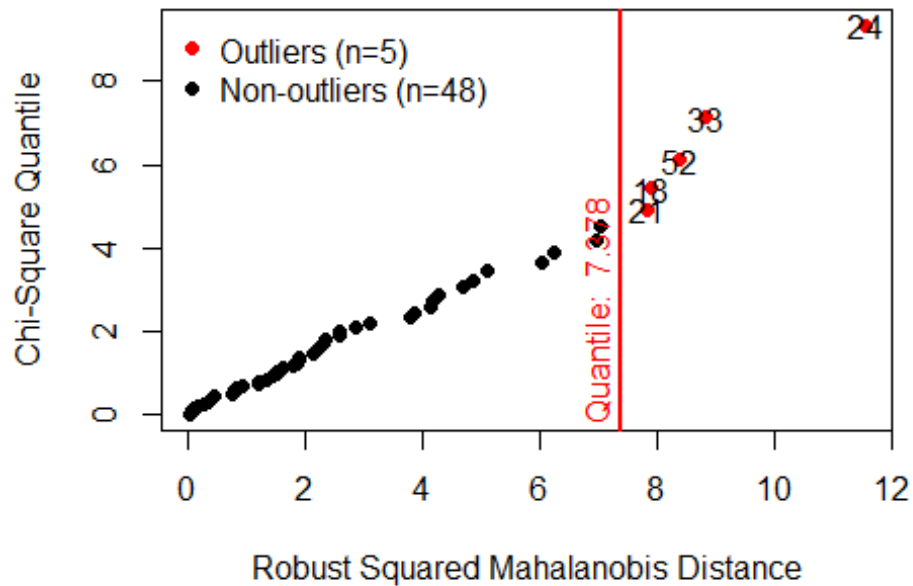
```
library(MVN)
result <- mvn(normD, mvnTest = "hz", multivariatePlot = "contour")
```



Datos atípicos o influyentes en la normal multivariada (distancia de Mahalanobis y gráfico QQplot multivariado)

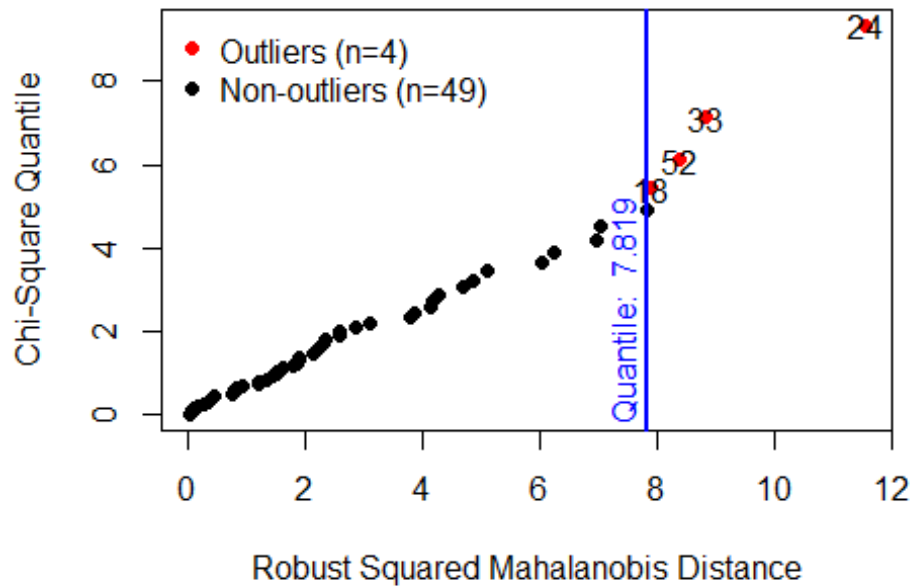
```
maha <- mvn(data = normD, mvnTest = "hz", multivariateOutlierMethod = "quan")
```

Chi-Square Q-Q Plot



```
adjustMaha <-  
mvn(data=normD,mvnTest="hz",multivariateOutlierMethod="adj")
```

Adjusted Chi-Square Q-Q Plot



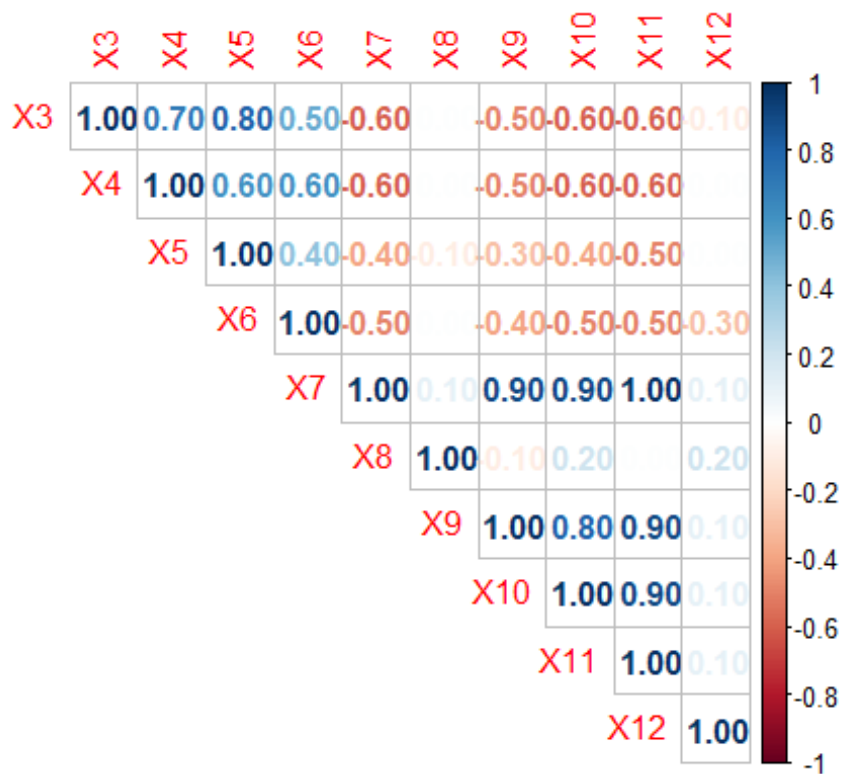
De Las gráficas obtenidas, La distancia de Mahalanobis declara 5 observaciones como valor atípico multivariado, mientras que La distancia Mahalanobis ajustada declara 4

Matriz de correlaciones

```
library(corrplot)

## corrplot 0.92 loaded

correlacion<-round(cor(D), 1)
corrplot(correlacion, method="number", type="upper")
```



Análisis de componentes principales

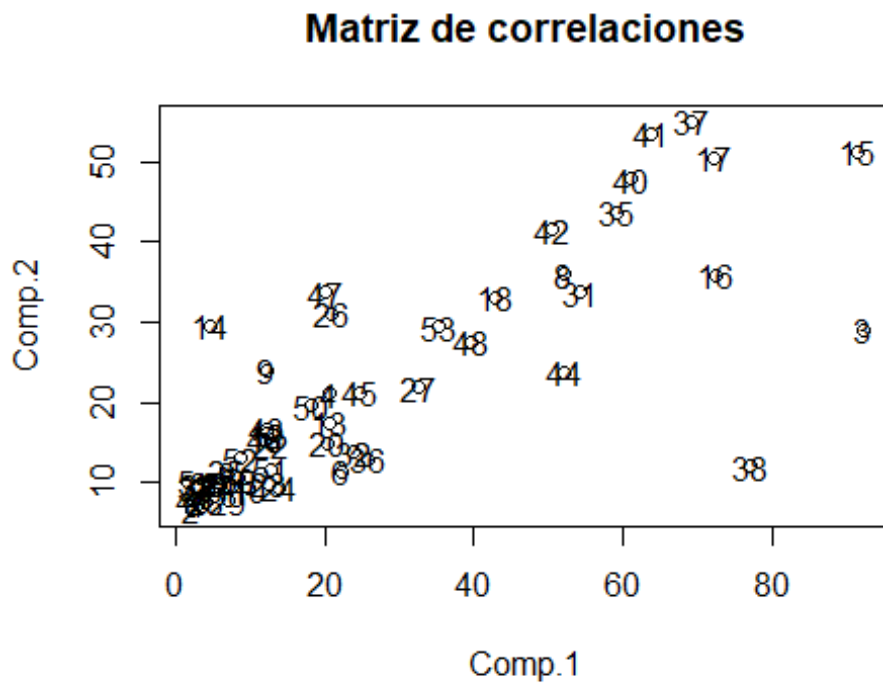
```
library(stats)
library(factoextra)

## Loading required package: ggplot2

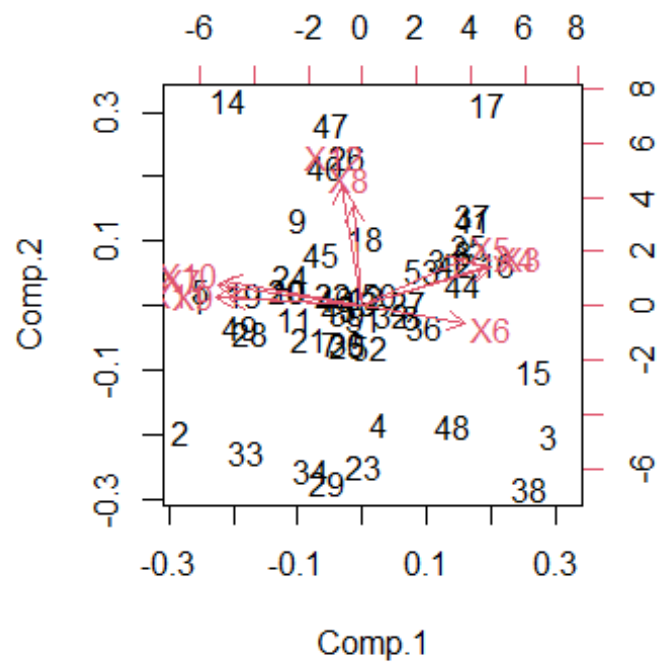
## Welcome! Want to learn more? See two factoextra-related books at
https://goo.gl/ve3WBa

library(ggplot2)
datos=D
cpS=princomp(datos, cor=TRUE)
cpaS=as.matrix(datos)%*%cpS$loadings
```

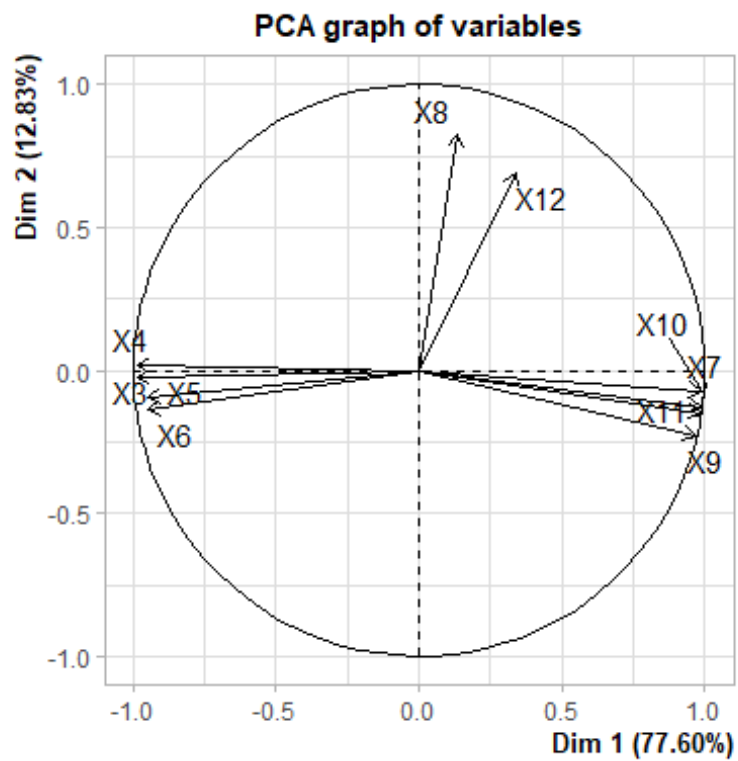
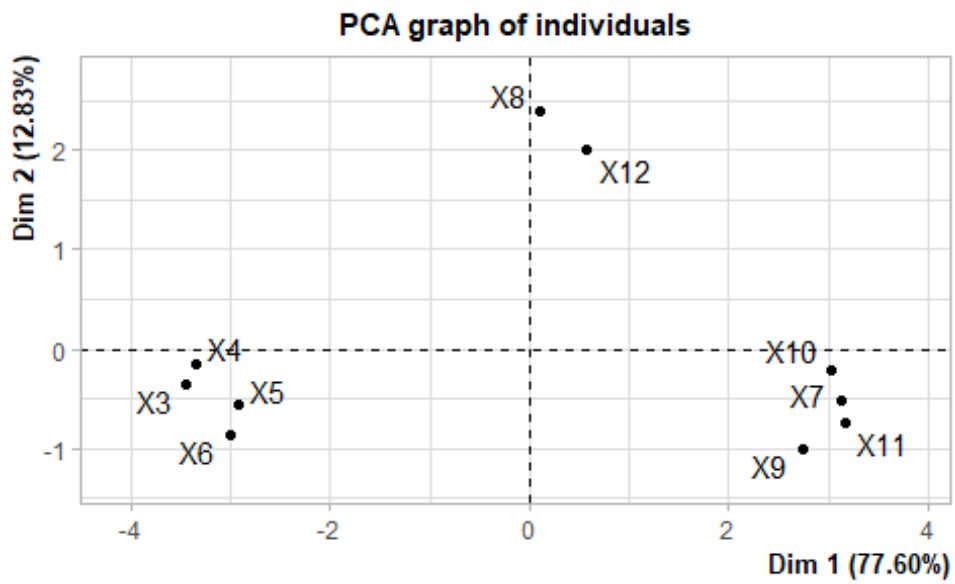
```
plot(cpaS[,1:2],type="p", main = "Matriz de correlaciones")
text(cpaS[,1],cpaS[,2],1:nrow(cpaS))
```



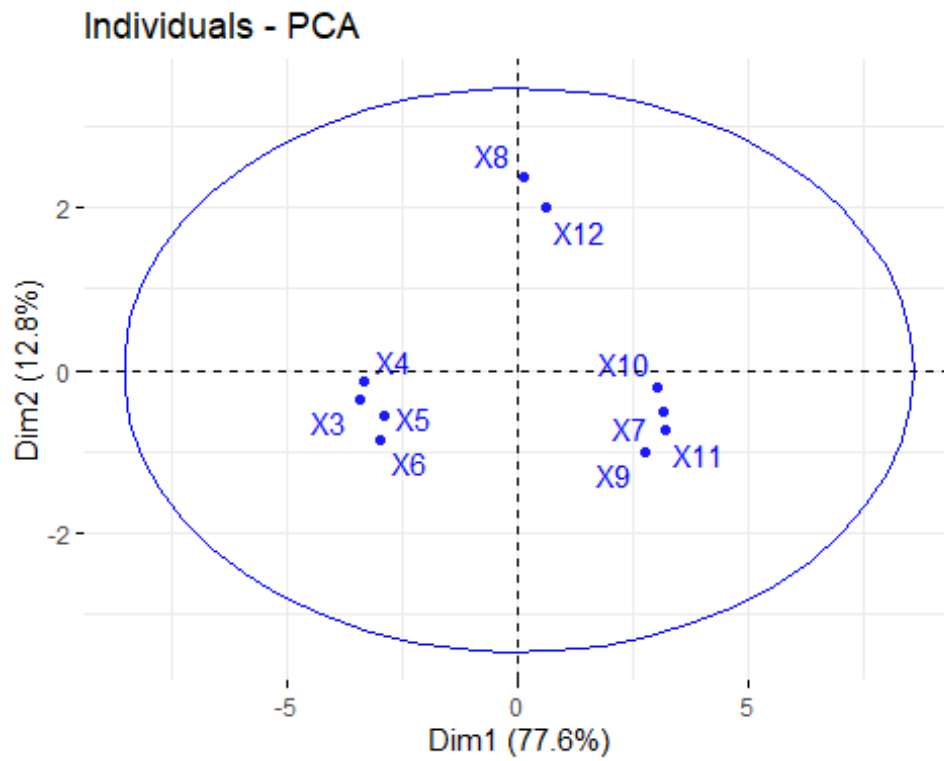
```
biplot(cps)
```



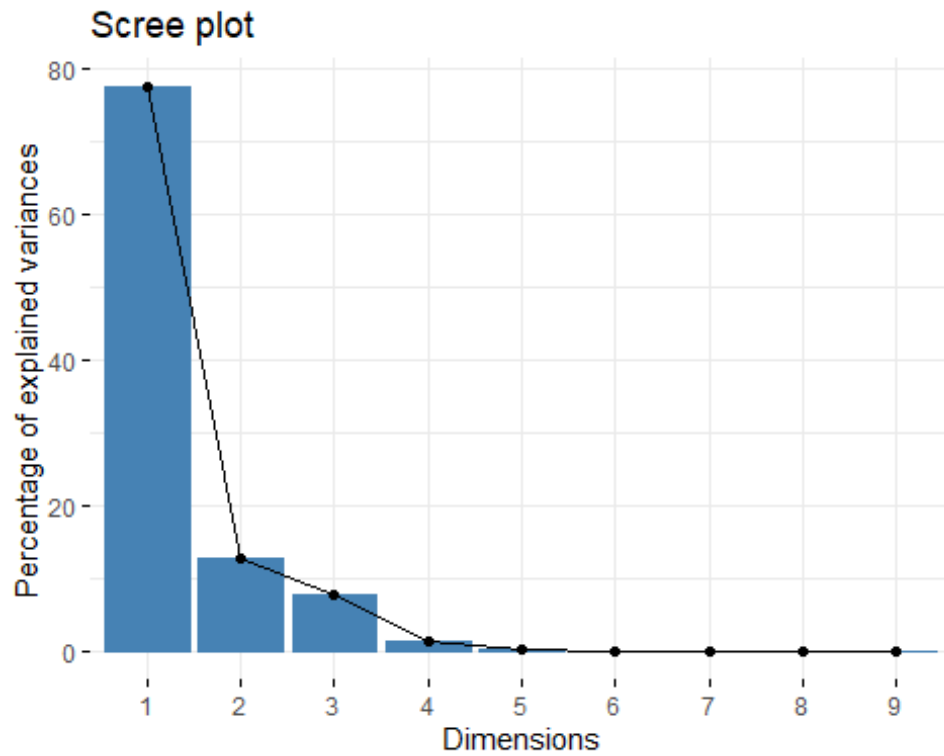

```
library(FactoMineR)
library(factoextra)
library(ggplot2)
datos=correlacion
cp3 = PCA(datos)
```



```
fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```



```
fviz_screepLOT(cp3)
```



```
fviz_contrib(cp3, choice = c("var"))
```

Contribution of variables to Dim-1

